Gu Test: A Progressive Measurement Of Generic Artificial Intelligence

Author: Lifan Gu□□□□□

Turing Test is subjective [1]. It is an empirical test, not a scientific experiment. Langauge complexity is much less than the human intelligence complexity. So Turing Test is invalid.

Sciences are different from mathematics. Scientific experiements only can falsify, but never prove unlimited possiblities. Scientific research is an ongoing process, should always open to new experiments.

So existing empirical tests for artificial intelligence (AI) technologies, such as the regular Go games played by AlphaGo Zero and other computer Go systems, the simulations and road tests for self-driving cars, the datasets for natural langauge understanding, etc. are also inadequate.

Technological Singularity is baseless. Driverless cars with no constrains (i.e. SAE level 5 automated-driving) are impossible. There are problems in the definition of SAE level 4.

In reality, there is no way to prove a car with SAE level 4 automated-driving ability, especially when the mode evolution in future is not stable. So new concepts of AI and new definitions of automated-driving should be studied.

In this paper, I will discuss the problems in Turing Test, the problems in existing testing of AlphoGo Zero, self-driving cars [2], natural langauge understanding, and the problems in the mainstream textbook AI: A Modern Approach. Then I will introduce a concept of integrity to solve some critical problems in intelligence studies, and propose Gu Test, a progressive measurement of generic artificial intelligence, based on falsifiability. These could help to develop scientific intelligence theories gradually.

1. The Problems in Turing Test

Turing Test is invalid, but still cause misleading widely in AI research so far. Many existing tests for AI techonologies have similar problems as Turing Test. So it is important to analyze its problems and clarify the misleading.

Turing Test is subjective. Testing it with different people could yield very different results. People with different knowledges, especially with different understanding levels of computer technologies, could give very different results. The subjectiveness of Turing Test cause unstable results, which makes Turing Test invalid.

Moreove, language complexity is much less than human intelligence complexity. Humans have much more intelligence than language level intelligence [3]. So Turing Test is not valid by making judgement of intelligence based on language conversation. Indistinguishablity between humans and computers by language conversations does not mean equivalence of intelligence.

Turing Test is also an empirical test, not a scientific experiment.

Sciences are different from mathematics. Scientific experiements only can falsify, but can never prove unlimited possiblities. Actually, equivalence of intelligence between humans and computers can never be proved, but only can be falsified.

Scientific research is an ongoing process, should always open to new experiments. If computers pass some tests, other people still could design new tests to disprove.

Scientific experiments should be done with strictly controlled conditions, to test the underlying principles. Scientific conclusions can only be derived from these principles based on the strict conditions. From empirical tests, people can not derive scientific conclusion.

Other existing tests for AI technologies have many similar problems. In the next sections, I will discuss the testing problems for computer Go systems, self-driving cars, and natural langauge understanding.

2. AlphaGo Zero's Superhuman Claim

The AlphaGo Zero paper on Nature magazine [4] claimed superhuman performance. However it did not provide any evidences for this claim, did

not provide any evidences to show AlphaGo Zero is superior to GENERIC human even in Go gaming.

AlphaGo Zero defeated AlphaGo Master and other AlphaGo series programs is not an evidence of superhuman, because these computer Go systems also suffer from the limitations of AI.

Even if AlphaGo Zero could defeat all human players in regular Go games, this still does not provide superhuman evidences, because these human players still do not know the limitations of AlphaGo Zero and AI yet.

As long as someone could design some scientific experiments with strictly controlled conditions, to let human players know the weakness of AlphaGo Zero and AI [5], human players still could defeat AlphaGo Zero and other computer Go systems in fair Go games.

Scientific experiments are different from regular gaming. Regular gaming is just to win a game or some games. Scientific experiments is to falsify some assumptions or let the assumptions pass the experiments. Scientific research should always open to new experiments.

Scientific experiments should be done with strictly controlled conditions. The conclusions of the experiments only can be derived from the controlled conditions and the results.

Scientific experiment results should open to discussions, so other people could discuss whether the experiments are valid and whether the interpretations of these results are justified or not.

The Go games played by AlphaGo Zero and other computer Go systems are just regular GAMES or empirical tests, not scientific experiments. People cannot derive scientific conclusions from these games or empirical tests.

So the superhuman claim from the AlphaGo Zero paper on Nature magazine is not a scientific conclusion. Actually, as already analyzed, this superhuman claim is baseless.

Go is a game with simple rules and good abstraction, but still significant complexity, which make it an ideal tool to design scientific experiments [6].

It is much easier to isolate various factors and figure out the intelligence principles in such simpler experiments with strictly controlled conditions.

3. Test Automated Driving

The current simulations and road tests for self-driving cars are not scientific experiments. They are just empirical tests without strictly controlled conditions. So from these simulations and road tests, we cannot derived any scientific conclusions [7].

Driverless cars with no constrains (i.e. SAE level 5 automated-driving) are impossible, which could be verified by scientifc experiments with the AI technologies used by these self-driving cars.

There are problems in the definition of SAE level 4. In different areas, there are very different requirements for SAE level 4 automated-driving techniques. Even in the same areas, if the mode evolution in future is not stable, severe problems could appear with large probabilities even if the cars already passed SAE level 4 in previous modes.

Scientifc experiments with the AI technologies used by self-driving cars could verify that when the mode evolution is not stable, severe problems with large probabilities could occur in the systems with very little problems in previous modes. So in reality, there is no way to prove a car has truly passed SAE level 4 automated-driving.

Although people could solve some problems in self-driving cars case by case, if they do not know the underlying principles, they would not solve certain important root causes of the problems. These root causes could appear in very different forms in future, especially if the mode evolution is not stable [8].

To understand unstable mode evolution, we need understand the underlying principles of intelligence in these AI technologies first. So it is a luck that we could study the AI problems of with theoretic analyses, and design experiments based on these analyses, to verify the AI problems with simpler systems, such as AlphaGo Zero.

I wrote before: Go gaming is strictly defined within a very small space. Industrial automations are typically designed in environments well controlled, but not strictly defined. Car driving is regulated, but the environment is not well controlled.

Many industrial automation problems can not be solved yet. As said, there are still serious problems even in AlphaGo Zero which could be verified by scientific experiments.

Based on my previous analyses and my experiment plans, I have reasons to believe: many years after large-scale deployment of self-driving cars, regular people would have enough chance to interact with self-driving cars from different aspects and trigger unstable mode evolutions which cause severe problems with large probabilities. However, at that time, it would be too late.

The technologies for traditional automobiles, such as electronics, powertrain, and other mechanics, etc. are based on concrete sciences whose main principles are already well tested in sciences. However, AI for automated driving is empirical, with no scientific foundation. It could be very unstable in future mode evolution. So testing automated driving vehicles in a similar way to testing traditional vehicles is very misleading.

I do not study automated driving directly myself. It is better to do foundamental studies first and figure out the underlying principles of intelligence. By verifying these principles and the problems in AI with simpler and undangerous AI systems, we could avoid the problems of deploying unmature self-driving cars in large scale.

4. The Problems in AI: A Modern Approach

There are problems in the philosophical foundation of the mainstream textbook: the 3rd edition of AI: A Modern Approach [9], which directly cause the problems of testing theories and methods in this textbook.

In the 1st edition of this book, there is an introduction of Socrates, with a reference: "Socrates asks Euthyphro, 'I want to know what is characteristic of piety which makes all actions pious ...'". Socrates was actually talking about a mode of mind: piety, an important topc of intelligence studies. Unfortunately, these introduction dissapeared in the 3rd edition of this book.

Galileo set Socratic method and experiments as the foundation of sciences [10]. By removing the introduction of Socrates, the textbook removed one pillar of sciences. The experiment method, another pillar of sciences, does not exist in the textbook due to its Aristotle thinking mode [11], the relevance of Turing Test, and Wind tunnel approach. So the textbook does not have scientific foundation. These are also the main problems in current AI sector [12].

I will discuss these problems one by one.

The textbook states: "Aristotle... was the first to formulate a precise set of laws governing the rational part of the mind", which is not true in physical sciences, biology, and mathematics.

The attitude of Aristotle thinking mode is like this: that's it, you have to believe me, there is no need to do experiments (although they are human, not God). This is exactly what Galileo criticized. Although in different fields, they could derive very different assertions.

As said, the rationale of sciences is based on Socratic method and experiments. Lacking of these elements, Aristotle thinking mode does not have integrity in physical sciences. Actually Galileo and Boyle formally denounced Aristotle thinking mode.

Robert Boyle further suggested the essences of matters rely on their internal compositions and structures, and should not be confused by their external characteristics [13]. Without understanding the internal structures and complexities of intellignece, AI has the similar problems of Aristotle thinking mode.

Aristotle's biological classification is static, and based on external characteristics, too. Darwin's evolution theory suggested species are dynamic and in evolution, which implies certain internal ambiguity in biological classification. Modern biology tries to improve biological classification by gene and other studies, which actually causes more ambiguity, even spread to the top level classification. So the ambiguity in biological classification is fundamental, and cannot be eliminated by logic. Aristotle thinking mode does not have integrity in biology.

In mathematics, Gödel even proved the problems of Aristotle's syllogisms. The rationale of mathematics is actually different from Aristotle thinking mode, and needs more intelligence components and elements, which should be studied further in depth.

So the rationales of physical sciences, biology, mathematics, actually conflicts with Aristotle thinking mode, which could not be "a precise set of laws governing the rational part of the mind" [14]. Lacking of integrity, Aristotle thinking mode could cause severe problems in AI.

The textbook also states: "Turing deserves credit for designing a test that remains relevant 60 years later", which is obviously not true. As analyzed in section 1, Turing Test cause severe misleading. To understand intelligence and test AI correctly, we need clarify such misleading.

The textbook promotes a Wind Tunnel approach. However, the designs of workable wind tunnels, engines, and airplanes all depends on physical sciences. The status of physical sciences in Wright Brothers' age is very different from the status of intelligence studies today.

Forces could be understood correctly only after Galileo made critical abstraction over stillness and movement. George Cayley could figure out the underlying principles and forces of flight only after Newton formed the systematic theory of forces. Without these researches, even with a wind tunnel, Aristotle or even Leonardo da Vinci could not design an airplane successfully.

However, in intelligence studies, we still do not know the fundamental principles, to distinguish the problems of Aristotle thinking mode and make scientifc breakthrough; we still do not know the reasons and complexities of integrity which is essential to scientifc development; we still do not know how to analyze the unstable mode evolution in future which is much more critical in intelligence than in physical sciences.

So Wind Tunnel approach would not work for intelligence studies now. We need structural and systematic analyses of human intelligence first. The studies of language intelligence could provide many important insights.

5. Measure Language Intelligence

AI could do searches well and have a much better memory for text contents than humans. AI even could achieve many progresses in machine translation. However, AI does not really understand semantics. There is a Chinese room issue, which could be verified.

AI could not process high-order logic properly, could not recognize sophism, could not recognize wrong thinking modes, such as Aristotle thinking mode.

So replying on AI to make judgement could cause severe problems in juridical practice, scientific researches, education, medical practice, etc. Asking students to obey computer's thinking mode could damage their intelligence development.

The current testing datasets for language understanding, such as SQuAD, CoQA, QuAC, NLVR^2, GLUE series, cannot measure the real difference between human and Natural Language Processing (NLP). They cannot help much on high-order logic processing, recognizing sophism, verify Chinese room issues, etc.

All of these datasets fall into the traps of Aristotle thinking mode. They can not recognize wrong thinking modes, and are not scientific methods.

To understand human intelligence, we need a structural and systematic

analysis of human intelligence. I defined certain main intelligence levels: language level, philosophical level, mathematical level, scientific level, all with different requirements and criteria.

Langauge intelligence is an important characteristic of human intelligence. Other known lives do not have advanced language ability. Langauge is also an important media for human knowledge, the basis for philosophy, mathematics, sciences, etc.

Based on languages, humans developed two important branches of studies: mathematics and philosophy. Mathematics develops towards accuracy. Philosophy develops towards integrity.

Sciences originates from philosophy, so sciences also develop towards integrity. More than philosophy, sciences make conclusions based on experiments of falsifiability with strictly controlled conditions. Beyond philosophy, sciences also gradually introduce accuracy and mathematics.

Mathematics does not meet the criteria of sciences. It even does not have integrity [15].

Based on these structural and systematic studies of human intelligence, people could measure langauge intelligence much better.

6. Gu Test

I introduce a concept of integrity, to clarify some misunderstanding and solve certain critical problems in brain and intelligence studies.

Lives at gene or animial level do not have integrity [16]. So there are severe problems in The Selfish Gene (or The Immortal Gene) theory.

Regular natural language expressions or conversations do not have integrity. Debates themselves do not have integrity. Sophism causes severe problems.

Gödel even proved mathematics does not have integrity. Turing Machine is a subset of mathematics, which also does not have integrity.

The name "universal approximation theorem" could be misleading, because there is a strict limitation of what it could approximate. Even the simple three-body problem in physics could cause issues. Brain and intelligence studies are much more complicated. Artificial neural networks with sensors do not have integrity.

However, sciences develop with integrity as the goal. The sciences of intelligence need study how integrity really develops.

I am designing a procedure to measure human specific intelligence progressively in AI technologies [17] and discover their problems based on falsifiability, to avoid the problems in empirical tests, and study what conditions could trigger unstable mode evolutions and significantly increase the probabilities of these AI problems to dangerous levels.

The test procedure also could provide some insights how integrity develops and how the fundamental principles of intelligence work under strictly controlled conditions.

Based on my studies, I propose:

1) A 4-dimension experiment space to test the intelligence of computer Go systems in Go games and discover the problems in the AI technologies used by these systems, especially to test AlphaGo Zero's superhuman claim or any such implications.

Since there is only one opportunity to gather certain important experiment results before computer Go systems could be adjusted by humans, the first round experiment should be done on the strongest Computer Go system with large-scale experiments [18].

I choice to start experiments with computer Go systems, because some important factors could be well isolated in such simpler systems with good abstraction.

2) Research and experiment schemes for languages, to study the expressing power and limitations of various languages, including natural languages, mathematics, music, etc., and study personalities, thinking modes, and mode transitions behind language expressing.

To understand human intelligence, we need study the development of commonsense, concepts, principles, theories, etc. We need study the fundamental principles in the development of natural lanugages, philosophy, mathematics, and sciences, especially the development of integrity.

To understand the problems in AI technologies, we need study high-order logic processing, sophism, Chinese room issues, etc.; and study the problems in computer thinking mode and Aristotle thinking mode.

Existing testing dataset series, such as SQuAD, CoQA, QuAC, GLUE,

NLVR^2(Natural Language for Visual Reasoning for Real), etc., do not help much on these issues.

3) Plans to study the relations between brains, mind, and human specific intelligence, started from the problems in The Selfish Gene (or The Immortal Gene) theory.

Certain experiments could be designed to verify some problems and principles. The experiments will not be related to humans. First, need a survey of current brain researches.

Current progresses in neurosciences are mainly at physiological or animal level, such as vision, audio, motion, emotion, etc., which do not illustrate the essentials of human specific intelligence, and can not show how humans develop integrity based on life entities without integrity.

......

These studies and experiments require certain amount of resources. They are all fundamental researches for peaceful purposes with no profit prospect. They are scientific researches of the concepts, principles, theories, philosophies of intelligence, rather than personal or social tactics or maneuvers, etc.

Scientific disputes can only be resolved by experiments with strictly controlled conditions. No person could play the role of judge in sciences. Questions and negations to my opinions should be subject to open discussions and strict experiments before conclusions being made.

7. Future Researches

The studies and experiments could be extended to other AI technologies and systems, and other aspects of human specific intelligence, etc., in future.

More studies should be done to illustrate the reasons and complexity of integrity development structurely and systematically.

I study these critical problems for the welfare of all humans. However, my health degrades very quickly, some degrading could be irreversible. I cannot do further researches unless in safety personally and economically.

Sophism, misleading, and wrong interpretation of technologies and empirical test results, could damage scientific researches in future.

----

[1] Scientific research should be objective. The scientific principles in quantum physics are still objective, although quantum physics does introduce uncertainty. How to develop objective principles based on the uncertainty in quantum physics is a very important topic of scientific philosophy.

[2] I discuss two applications here: AlphaGo Zero and self-driving cars, because the superhuman claim of AlphaGo Zero was published on an important academic magazine Nature, and self-driving cars were widely advocated in many years (called driverless cars before) and relate to public safety. Several years ago, I already heard that the technologies of driverless cars was already ready, just the laws were behind, which obviously is not true.

[3] In section 5., I will discuss more on different intelligence levels.

[4] Mastering the game of Go without human knowledge, Published: 18 October 2017:

https://www.nature.com/articles/nature24270

[5] Actually I designed such experiments as introduced in the section 6 of this article, and requested Deepmind to do the experiments, but they have not accepted the experiments.

Scientific research should be based on open discussion and fair experiment. So the superhuman claim for AlphaGo Zero is not a scientific conclusion.

[6] I began to consider to use computer Go systems to measure AI technologies long before Deepmind started AlphaGo project.

[7] According to some news, in 2015 a blind man was allowed to take a driverless car alone, before the accident on 02/14/2016. Although the damage of this accident is minor, wrong judgment of driverless cars is very dangerous potentially, especially if the mode evolution in future is unstable.

"Steve Mahan, who is legally blind, was the first non-Google employee to

ride alone in the company's gumdrop-shaped autonomous car. The ride was in October 2015 in Austin. (Courtesy Waymo)", https://www.washingtonpost.com/local/trafficandcommuting/blind-man-sets-out-alone-in-googles-driverless-car/2016/12/13/f523ef42-c13d-11e6-8422-eac61c0ef74d_story.html,

"Steve Mahan, who is legally blind, takes what Waymo called the world's first fully autonomous ride in Austin in 2015, in an image provided by the Alphabet unit.", https://www.marketwatch.com/story/google-says-driverless-cars-are-ready-to-make-money-but-we-wont-know-if-they-do-2016-12-13.

[8] In a MIT lecture published on Feb 12, 2019, Drago Anguelov. a Principal Scientist at Waymo, admitted that there is a long tail of problems in self-driving cars: https://www.youtube.com/watch?v=Q0nGo2-y0xY.

The real situation could be more complicated than a long tail. The underlying principles of intelligence could help us to understand how the problems evolve, and transform, etc.

[9] The 3rd edition of AI: A Modern Approach is referred simply as "the textbook" in this section for convenience.

[10] Dialogue Concerning the Two Chief World Systems, Galileo Galilei (1632).

[11] Obviously, Aristotle thinking mode does not have integrity: Aristotle "counsels Alexander to be 'a leader to the Greeks and a despot to the barbarians, to look after the former as after friends and relatives, and to deal with the latter as with beasts or plants'", Alexander of Macedon, Green, Peter (1991), University of California Press. ISBN 978-0-520-27586-7. https://en.wikipedia.org/wiki/Aristotle.

However, in this article, I only discuss the problems of Aristotle thinking mode in academic research, specifically in physics, chemistry, biology, and mathematics. Integrity is an important concept I introduce in intelligence studies, to clarify certain misunderstandings and solve some problems.

[12] I inquired Waymo about some recent news of unreasonable behaviors of their cars, and asked them either to deny or to confirm the reports. I also requested experiments with AlphaGo Zero to verify some generic problems in AI technologies. They have not replied.

Of course, they have the right not to reply. However, by neither denying nor confirming the unreasonable behaviors of their cars, they actually declare

they do not follow scientific ways, which raises serious concerns because self-driving cars relate to public safety.

If the news reports are true, there could be serious problems in the cars. If the reports are not true, not clarifying them also could be dangerous.

Reverse usage of "wolf is coming" is dangerous, too. If faked news of problems (false warning of "wolf is coming") appeared several times without being clarified, when real problems appears (wolf is really coming) people would ignore it.

I do not make conclusions based on news or empirical results. However, I do urge Waymo and Deepmind to clarify the issues by open discussion and fair experiments.

Sciences require open discussion and experiments with strictly controlled conditions. Scientific conclusion only could be derived from strictly controlled conditions. Open discussion is to assure the experiments are valid and the interpretation of the experiment results is correct.

[13] The Sceptical Chymist, Robert Boyle (1661).

[14] Bertrand Russell even wrote: "Ever since the beginning of the seventeenth century, almost every serious intellectual advance has had to begin with an attack on some Aristotelian doctrine; in logic, this is still true at the present day", A History of Western Philosophy (1945).

The beginning of the seventeenth century, is exactly when scientific revolution started. Since then, civilization experienced a fast development. If open discussion and experiments, the two pillars of sciences, are removed in studies now, development could slow down, or even stop.

[15] For more details, please see my article: A Structural and Systematic Analysis of Human Knowledge and Studies.

[16] Gene or life entities do not have integrity, although life sciences have integrity as the goal of development. There are differences between lives and life sciences.

[17] Gu Test does not intend to distinguish humans from humans. It only measures the difference between generic human and machines, or between generic human and other animals. However, it could help humans to avoid the problems of computer thinking mode and Aristotle thinking mode, etc. I treat thinking mode differently from persons or personality.

[18] To controll the experiment conditions strictly, I suggest Deepmind

submit all the necessary hardware, software, data files, etc., to Congress Library, so all the versions could be verified before each experiment. Then I could submit my experiment plans.

Other people also could design their experiments based on such a service. Actually I do expect other people could design better experiments after my experiments have been done. Such a public service is better provided by the government.